

IMG Workspace User Guide

1. Introduction

IMG users can save selected genomes, genes, scaffolds, and functions as Workspace genome set, gene set, scaffold set, and function set, respectively. Workspace provides more functionalities compared with Analysis Cart:

- A user can have multiple named Workspace genome sets (gene sets, scaffold sets, function sets), while there is only one Genome Cart (Gene Cart, Scaffold Cart, Function Cart). This allows users to perform set comparisons and analyses over multiple data sets. For example, a user can compare genome sets using intersection or set difference, which is not possible using Analysis Cart.
- Workspace data sets are permanently stored until they are deleted by owners, while analysis carts are transient (i.e., lost after user logs out).
- Users can save a large number of genomes (genes, scaffolds, functions) in a Workspace data set, while analysis carts can only hold a much smaller numbers of objects (e.g., up to 20,000 genes).
- Workspace supports computation jobs using message system. Users can submit a large computation job that requires many hours of computation, and later return to IMG to check the results. These types of computations will time out in Analysis Cart.
- **(New Feature)** If a user belongs to at least one IMG group, he/she can share his/her own Workspace data sets with other group members. The user can also view and use Workspace data sets shared by other group members. This new feature allows interactive collaboration of group members working on a research topic. In order to use this new feature, users should first create their own IMG groups or ask to join other existing IMG groups. IMG group is a feature in MyIMG; there is another user guide in MyIMG main page.

2. My Workspace

The "Workspace" submenu under MyIMG will lead to a page listing counts of all the Workspace data sets a user owns and all computation jobs submitted by the user:

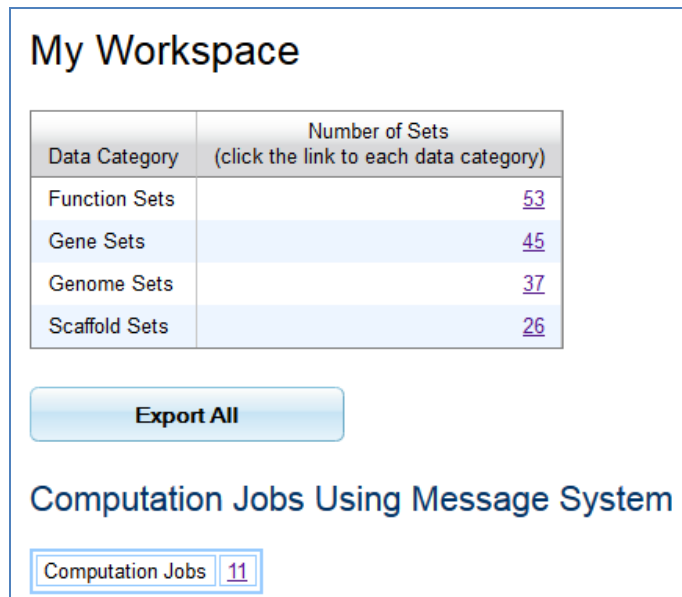


Figure 2.1: My Workspace Main Page

Clicking on each count will lead to the main page of Workspace Function Sets, Gene Sets, etc. as described in the following sections.

3. Workspace Function Sets

Workspace function sets main page (Figure 3.1) lists all the function sets created and owned by a user. If the user belongs to one or more IMG groups, then the table will also include function sets shared by other group members, and the table will have two additional columns:

- Owner: the owner of function set
- Shared with Group: all IMG groups that have access to this function set

Owners can share function sets with any IMG groups they belong to. They can also revoke the sharing. Two functions "Share Selected Set(s) with Group" and "Remove Sharing of Selected Set(s)" are provided for sharing and revoking, respectively. IMG Group dropdown list includes all IMG groups a user belongs to.

In order to share function sets, a user should:

1. select all function sets to be shared;
2. select the group to share in IMG Group dropdown list;
3. click the "Share Selected Set(s) with Group" button.

To revoke sharing, a user should:

1. select all function sets the user wishes to revoke sharing;
2. select the group to revoke sharing in IMG Group dropdown list;
3. click "Remove Sharing of Selected Set(s)" button.

My Workspace - Function Sets

Function Sets Import & Export Scaffold Search Gene Profile Set Operation

Add Selected to Function Cart Select All Clear All Remove Selected

Select	File Name	Number of Functions (click the link to each individual set)	Owner	Shared with Group
<input type="checkbox"/>	3_hydroxypropionate	3	me	-
<input type="checkbox"/>	3_hydroxypropionate_3_hydroxybutyrate	5	me	GBP
<input type="checkbox"/>	4_hydroxybutyrate	3	me	GBP
<input type="checkbox"/>	5pfam	5	me	GBP
<input type="checkbox"/>	Leucine_cog	10	me	GBP
<input type="checkbox"/>	Methanogenic	82	me	-
<input type="checkbox"/>	Nitrogen_fixer_func	17	me	-
<input type="checkbox"/>	Plastoquinol	4	me	-
<input type="checkbox"/>	Streptomycin_EC	12	me	GBP
<input type="checkbox"/>	jh_all_funcs_1a	12	me	GBP
<input type="checkbox"/>	scog	44	me	-
<input type="checkbox"/>	vancomycin	3	me	-
<input type="checkbox"/>	vancomycin_pfam	31	me	GBP

(Note: You can only share or delete your own datasets.)

IMG Group: [GBP](#) Share Selected Set(s) with Group Remove Sharing of Selected Set(s)

Figure 3.1 My Workspace - Function Sets

There are 4 buttons right above the table:

- "Add Selected to Function Cart": load functions in all selected function sets to the Function Cart.
- "Select All": select all function sets
- "Clear All": clear all selections
- "Remove Selected": delete all selected function sets (permanent deletion after user confirmation)

Workspace function set has 4 additional tabs with additional functions. We will describe these functions immediately below.

3.1 Import & Export

To import function sets into Workspace, browse and select a file with specified format, and click the "Import Function Sets" button.

To export function sets, first click and select function sets in the main page (Figure 3.1), and then click the "Export Function Sets" button.

My Workspace - Function Sets

Function Sets

Import & Export

Scaffold Search

Gene Profile

Set Operation

Import

You may import function sets from a file created by using the export feature below.
A file can also be successfully imported if it follows a [specific format](#).

File to upload:

Browse...

 No file selected.

Import Function Sets

Export

You may select one or more [your own](#) function sets from above to export. The exported file may be imported later into your workspace.
NOTE: Exported function sets contain IDs only. To export the contents of a function set, please go to the function set page.

Export Function Sets

Figure 3.2: Import & Export

3.2 Scaffold Search

Scaffold Search allows users to find whether any scaffolds of selected genomes have genes linking to functions in the selected function sets. To perform Scaffold Search, users:

- select one or more Workspace function sets;
- select one of more genomes in genome selection section;
- select Assembled/Unassembled/Both option for metagenome selection;
- click "Workspace Function-Scaffold Search" button.

Function Scaffold Search

Filter column: Scaffold ID Filter text Apply

Export Page 1 of 1 << first < prev 1 next > last >> All

Column Selector Select Page Deselect Page

Select	Scaffold ID	Scaffold Name	Genome	EC:2.6.1.56	EC:2.1.4.2	EC:3.1.3.39	EC:1.1.1.18	EC:2.7.7.33
<input type="checkbox"/>	641522711	Streptomyces griseus subsp. griseus NBRC 13350: NC_010572	641522653	0	2	1	2	0
<input type="checkbox"/>	2554308905	Streptomyces collinus Tu 365 : CP006259	2554235367	0	0	0	3	0
<input type="checkbox"/>	2554308906	Streptomyces collinus Tu 365 : CP006261	2554235367	0	0	0	0	0
<input type="checkbox"/>	2554308907	Streptomyces collinus Tu 365 : CP006260	2554235367	0	0	0	0	0
<input type="checkbox"/>	2554308905	Streptomyces fulvissimus	2554235367	0	0	0	4	0

Figure 3.3: Function Scaffold Search Result

Function Scaffold Search result (in Figure 3.3) shows the gene counts in each scaffold associated with individual functions in the function set(s). Users can click on non-zero gene counts to view the gene list (in Figure 3.4). Users can also select scaffolds from the list to save to Workspace scaffold set.

Scaffold Function Profile Gene List

Function: EC:1.1.1.18 (Inositol 2-dehydrogenase.)
Scaffold: 2554308905

Filter column: Gene ID Filter text Apply

Export Page 1 of 1 << first < prev 1 next > last >> All

Column Selector Select Page Deselect Page

Select	Gene ID	Function ID
<input type="checkbox"/>	2555756044	EC:1.1.1.18
<input type="checkbox"/>	2555761043	EC:1.1.1.18
<input type="checkbox"/>	2555761411	EC:1.1.1.18

Export Page 1 of 1 << first < prev 1 next > last >> All

Add Selected to Gene Cart Select All Clear All

Save Expand Display

Save Genes to My Workspace

Figure 3.4: Scaffold Function Profile Gene List

Users can select and save genes into Workspace gene set from the result profile gene list.

3.3 Gene Profile

Gene Profile function allows users to check whether any genes in a selected Workspace gene set is annotated with any functions in selected Workspace function set(s). To perform this function, users:

- select one or more Workspace function set(s);
- select a Workspace gene set in "Using genes in Gene Set" dropdown list;
- click the "Workspace Function-Gene Profile" button.

Function Set Gene Profile (646311909essen)

Genes with no selected function counts are not displayed.

Filter column: Gene ID Filter text Apply ?

Export Page 1 of 1 << first < prev **1** next > last >> All

Column Selector Select Page Deselect Page

Select	Gene ID	Gene Product Name	5pfam	Streptomycin_EC	vancomycin
<input type="checkbox"/>	646318234	phosphodiesterase	1	0	0

Export Page 1 of 1 << first < prev **1** next > last >> All

Add Selected to Gene Cart Select All Clear All

Expand Gene Table Display


hint: Limit gene selection and display options to avoid timeout.

Figure 3.5: Function Set Gene Profile

Users can select genes in the result table (Figure 3.5) to add to Gene Cart or to expand the gene table display with additional fields. Click on a non-zero count leads to a new page listing individual functions in the set associated with the selected gene (Figure 3.6). Users can select to save functions into Function Cart or a new Workspace function set.

Functions in Function Set for Selected Gene

Gene (646318234): phosphodiesterase. Function Set: 5pfam.

Filter column: Function ID Filter text: Apply 

Export Page 1 of 1 << first < prev 1 next > last >> All


Column Selector Select Page Deselect Page

Select	Function ID	Function Name
<input type="checkbox"/>	pfam00013	KH domain

Export Page 1 of 1 << first < prev 1 next > last >> All

Add Selected to Function Cart Select All Clear All

Save Functions to My Workspace



Even though you can save large amount of data into workspace, many profile functions will timeout for extremely large workspace datasets

Figure 3.6: Functions in Function Set for Selected Gene

3.4 Set Operation

There are 3 types of set operations on selected Workspace function sets (Figure 3.7):

- Union of 2 or more function sets: Select function sets, enter Save to File name (which will be a new Workspace function set), and click "Save Union to Workspace" button.
- Intersection of 2 or more function sets: Select function sets, enter Save to File name (which will be a new Workspace function set), and click "Save Intersection to Workspace" button.
- Difference of Two Sets: Select first and second function sets in the corresponding dropdown lists, enter Save to File name (which will be a new Workspace function set), and click "Save Difference to Workspace" button.

Function Sets
Import & Export
Scaffold Search
Gene Profile
Set Operation

Union or Intersection of Sets

You may select two or more function sets from the table above to get their union or intersection.

Save to File name:

(Special characters in file name will be removed and spaces converted to _)

Save Intersection to Workspace
Save Union to Workspace

Difference of Two Sets (First - Second)

First function set:

Second function set:

Save to File name:

(Special characters in file name will be removed and spaces converted to _)

Save Difference to Workspace

Figure 3.7: Set Operation

4. Workspace Genome Sets

Workspace genome sets main page (Figure 4.1) lists all the genome sets created and owned by a user. If the user belongs to one or more IMG groups, then the table will also include genome sets shared by other group members, and the table will have two additional columns:

- Owner: the owner of genome set
- Shared with Group: all IMG groups that have access to this genome set

Genome sets sharing is similar to function sets sharing. Check Section 3 to find out how to share Workspace data sets.

My Workspace - Genome Sets				
<div> Genome Sets Import & Export Function Profile Blast Set Operation </div>				
<div> Add Selected to Genome Cart Select All Clear All Remove Selected </div>				
Select	File Name	Number of Genomes (click the link to each individual set)	Owner	Shared with Group
<input type="checkbox"/>	10_pfam_genomes	254	me	-
<input type="checkbox"/>	AAA471	49	me	-
<input type="checkbox"/>	Archaea	469	me	-
<input type="checkbox"/>	Bacteria	20345	me	-
<input type="checkbox"/>	Chlorobium	6	me	GBP
<input type="checkbox"/>	Etoliko	2	me	-
<input type="checkbox"/>	Euk	121	me	-
<input type="checkbox"/>	GEBA	208	me	-
<input type="checkbox"/>	GEBA_relevance	210	me	-

Figure 4.1: My Workspace - Genome Sets

Workspace genome set has 4 additional tabs with additional functions. We will describe these functions immediately below.

4.1 Import & Export

Import & Export is similar to Workspace function set Import & Export described in Section 3.1.

4.2 Function Profile

Users can select to perform Genome Set Function Profile based on (i) a selected Workspace function set or (ii) a function type (COG, Pfam, etc.). To perform Genome Set Function Profile, users

- select one or more genome sets;
- select one of the following (See Figure 4.2):
 - i. select a Workspace function set in the "Use only functions in set" dropdown list, or
 - ii. select a function type in the "Use all functions of type" dropdown list;
- select Assmbled/Unassembled/Both option if there are any metagenomes in the selection;
- click "Genome Set Function Profile" button.

My Workspace - Genome Sets

Genome Sets

Import & Export

Function Profile

Blast

Set Operation

Genome Set Function Profile

Select no more than 50 genome sets to generate genome set vs. function profile.

hint:

Limit number of selected genome sets and/or number of functions to avoid timeout.

☒ Use only functions in set:

3_hydroxypropionate

☐ Use all functions of type:

COG

MER-FS Metagenome:

Assembled

Genome Set Function Profile

Submit as Computation Job Using Message System

You may submit a genome set function profile computation to run in the background.

☒ Save as a new job with name:

☐ Replace the selected job:

job04451

Submit Computation

Figure 4.2: Genome Set Function Profile

Genome Set Function Profile result page lists functions and number of genomes in each Workspace genome set having this function. For example, Figure 4.3 shows that genome set "Rhizobium" has one gene with KO:K09709 and "Streptomyces" has 13 genes with the same KO.

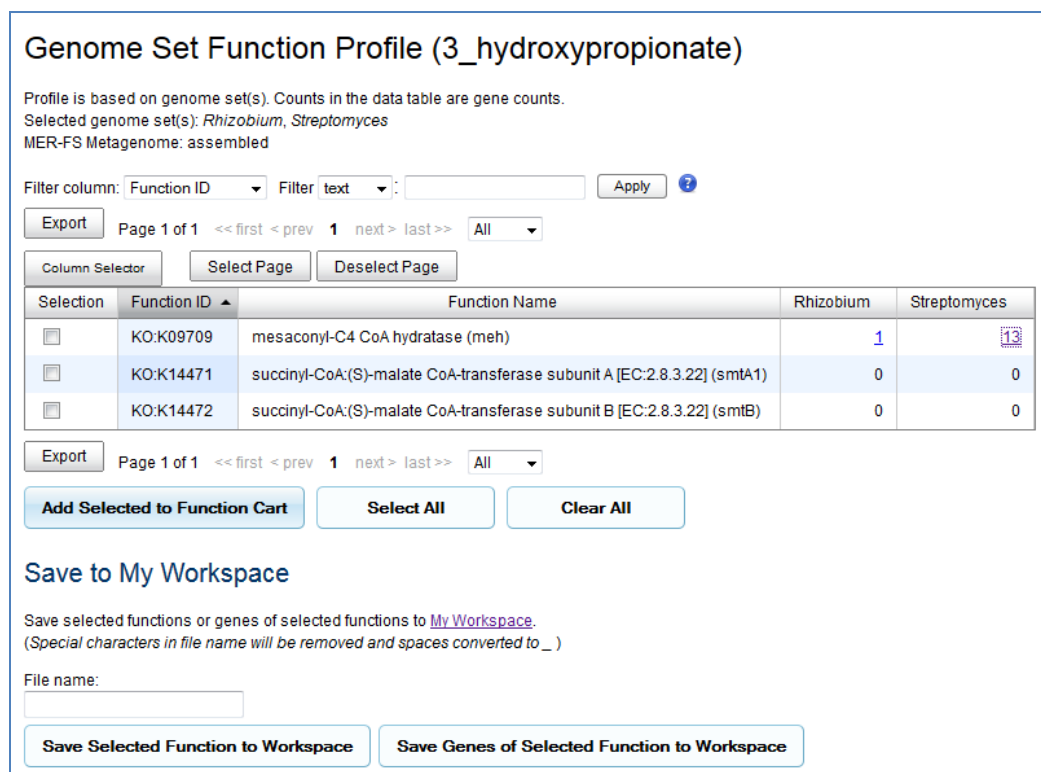


Figure 4.3: Genome Set Function Profile Result Page

Users can select functions to be loaded to Function Cart or saved to a new Workspace function set. Clicking on the genome count will lead to individual gene list (Figure 4.4):

Genome Set Function Profile Gene List

Function: KO:K09709 (mesaconyl-C4 CoA hydratase)
Genome Set: Streptomyces

Filter column: Filter

Page 1 of 1 << first < prev 1 next > last >>

Select	Gene ID	Gene Name	Genome
<input type="checkbox"/>	645399600	hypothetical protein	Streptomyces hygroscopicus ATCC 53653
<input type="checkbox"/>	645979150	mesaconyl-C4-CoA hydratase (EC 4.2.1.-)	Streptomyces sp. AA4
<input type="checkbox"/>	645979994	hypothetical protein	Streptomyces sp. AA4
<input type="checkbox"/>	646969843	hypothetical protein	Streptomyces bingchenqensis BCW-1
<input type="checkbox"/>	647984707	MaoC like domain-containing protein	Streptomyces sp. e14
<input type="checkbox"/>	648752919	conserved hypothetical protein	Streptomyces violaceusniger Tu 4113
<input type="checkbox"/>	651727112	hypothetical protein	Streptomyces nanchangensis NS3226
<input type="checkbox"/>	2513524224	Uncharacterized conserved protein	Amycolatopsis sp. 75iv2, ATCC 39116
<input type="checkbox"/>	2516057802	Uncharacterized conserved protein	Streptomyces scabrisporus DSM 41855
<input type="checkbox"/>	2519023898	Uncharacterized conserved protein	Streptomyces sulphureus DSM 40104
<input type="checkbox"/>	2522458049	Uncharacterized conserved protein	Streptomyces sp. WMMB 322
<input type="checkbox"/>	2522464010	Uncharacterized conserved protein	Streptomyces sp. WMMB 714
<input type="checkbox"/>	2526040245	Uncharacterized conserved protein	Streptomyces sp. TAA486

Page 1 of 1 << first < prev 1 next > last >>

Figure 4.4: Genome Set Function Profile Gene List

Users can select genes to be loaded into Gene Cart or saved to a new Workspace gene set.

If users select genome sets with large number of genomes and genes, Genome Set Function Profile will time out. To avoid this problem, users can submit background computation request instead. To do so, specify a job name, and click the "Submit Computation" button (instead of the "Genome Set Function Profile" button) shown in Figure 4.2. Computation job is described in more detail in Section 7.

4.3 Blast

Blast tab provides the usual blast function. Users enter or paste a sequence, select a blast program and an E-value cut-off, and then select a list of genomes to blast against. Click the "book icon" at the top (in Figure 4.5) to find more information on blast.

BLAST ?

Find matches in genomes selected below. [All isolate genomes](#) in IMG will be used if none are selected. This is equivalent to the old 'All IMG Genes - One large Database' option.
 If you select more than 100 genomes [All isolate genomes](#) in IMG will be used
 If genomes were selected in the Genome Browser, please use the ["Show All Genomes"](#) button below to access [All IMG genomes](#).

Sequence

Program: **blastp (Protein vs. Protein)**

E-value: **1e-5**

MER-FS Metagenome: **Assembled**

Genome Filter

Domains: (*)=Microbiome,
 (B)acteria, (A)rchaea, (E)ukarya, (P)lasmids, (G)ragment, (V)iruses.
 Genome Completion: [F]inished, [P]ermanent Draft, [D]raft.

Seq. Status: **All Finished, Permanent Draft and Draft** Domain: **Archaea**

☒ List ☐ Tree

ANME-1 cluster archaeon SCGC AAA252-L18 (SAK-pl.6 rarefaction # 69) (A)[D]
 AP Crenarchaeota archaeon JGI 005-G151416265346942 (A)[D]

Figure 4.5: Blast

4.4 Set Operation

Set operation is similar to set operations on Workspace function sets described in Section 3.4.

5. Workspace Gene Sets

Workspace gene sets main page (Figure 5.1) lists all the gene sets created and owned by a user. If the user belongs to one or more IMG groups, then the table will also include gene sets shared by other group members, and the table will have two additional columns:

- Owner: the owner of gene set
- Shared with Group: all IMG groups that have access to this gene set

Gene sets sharing is similar to function sets sharing. Check Section 3 to find out how to share Workspace data sets.

My Workspace - Gene Sets

hint: Select one or more gene sets to perform gene set analysis. Click on the gene set count to view and analyze genes in a particular gene set.

Gene Sets | Import & Export | Genomes & Scaffolds | Function Profile | Set Operation

Add Selected to Gene Cart | Select All | Clear All | Remove Selected

Select	File Name	Number of Genes (click the link to each individual set)	Owner	Shared with Group
<input type="checkbox"/>	4_hydrob	43	me	-
<input type="checkbox"/>	637000002essen	493	me	-
<input type="checkbox"/>	6370000290essen	110	me	-
<input type="checkbox"/>	637000126essen	635	me	-
<input type="checkbox"/>	637000184essen	306	me	-
<input type="checkbox"/>	646311909essen	221	me	-
<input type="checkbox"/>	TC_genes	311	me	-
<input type="checkbox"/>	atcc39243	7	me	-
<input type="checkbox"/>	bleomycin	32	me	-

Figure 5.1: My Workspace - Gene Sets

Workspace gene set has 4 additional tabs with additional functions. We will describe these functions immediately below.

5.1 Import & Export

Import & Export is similar to Workspace function set Import & Export described in Section 3.1.

5.2 Genomes & Scaffolds

The View Genomes or Scaffolds function allows users to view genomes or scaffolds of genes in selected gene set(s).

hint: Only scaffolds of **assembled** genes will be displayed or saved.

View Genomes or Scaffolds

You can view genomes or scaffolds of genes in the selected gene set(s).

View Genomes | View Scaffolds

Figure 5.2: View Genomes or Scaffolds

Users first select one or more Workspace gene sets and then click "View Genomes" or "View Scaffolds" buttons (in Figure 5.2) to view genomes or scaffolds of genes in selected gene sets, respectively. Figure 5.3 shows an example of scaffolds of selected gene sets. Users can select scaffolds in the result table to be loaded to Scaffold Cart or saved to a new Workspace scaffold set.

Scaffolds of Selected Gene Sets

Selected gene set(s): *griseola*

Filter column: Scaffold ID Filter text: Apply ?

Export Page 1 of 1 << first < prev 1 next > last >> All

Column Selector Select Page Deselect Page

Select	Scaffold ID	Scaffold Name	Genome Name
<input type="checkbox"/>	651716507	Kitasatospora griseola mevalonate pathway gene cluster (ggdps, mevk, mdpd, pmevk, ORF, hmgr, hmgs), complete cds.: AB037907	Kitasatospora griseola

Export Page 1 of 1 << first < prev 1 next > last >> All

Add Selected to Scaffold Cart Select All Clear All

Save Scaffolds to My Workspace

hint: Even though you can save large amount of data into workspace, many profile functions will timeout for extremely large workspace datasets

Save **selected scaffolds** to [My Workspace](#).
(Special characters in file name will be removed and spaces converted to _)

☒ Save to File name:
☐ Append to the following scaffold set:
☐ Replacing the following scaffold set:

Figure 5.3: Scaffolds of Selected Gene Sets

Users can also directly save the results to Workspace genome set or scaffold set using the "Save Genomes of Selected Gene Sets" and "Save Scaffolds of Selected Gene Sets" (in Figure 5.4), respectively.

Save Genomes to My Workspace

hint: Even though you can save large amount of data into workspace, many profile functions will timeout for extremely large workspace datasets

Save **all genomes of selected gene set(s)** to [My Workspace](#).
(Special characters in file name will be removed and spaces converted to _)

☒ Save to File name:
☐ Append to the following genome set:
☐ Replacing the following genome set:

Save Genomes of Selected Gene Sets

Save Scaffolds to My Workspace

hint: Even though you can save large amount of data into workspace, many profile functions will timeout for extremely large workspace datasets

Save **all scaffolds of selected gene set(s)** to [My Workspace](#).
(Special characters in file name will be removed and spaces converted to _)

☒ Save to File name:
☐ Append to the following scaffold set:
☐ Replacing the following scaffold set:

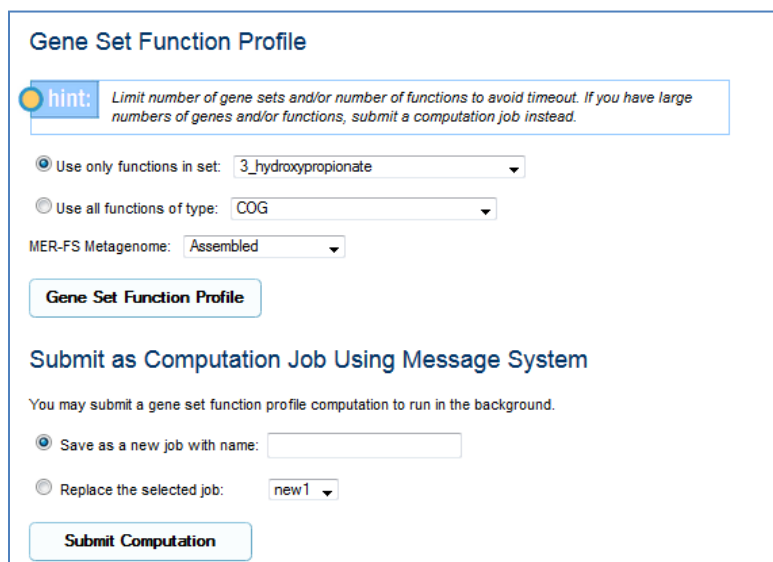
Save Scaffolds of Selected Gene Sets

Figure 5.4: Save Genomes or Scaffolds to My Workspace

5.3 Function Profile

Users can select to perform Gene Set Function Profile based on (i) a selected Workspace function set or (ii) a function type (COG, Pfam, etc.). To perform Gene Set Function Profile, users

- select one or more gene sets;
- select one of the following (See Figure 5.5):
 - i. select a Workspace function set in the "Use only functions in set" dropdown list, or
 - ii. select a function type in the "Use all functions of type" dropdown list;
- select Assembled/Unassembled/Both option if there are any metagenome genes in the selection;
- click "Gene Set Function Profile" button.



The screenshot shows a web form titled "Gene Set Function Profile". At the top, there is a hint box with a yellow circle icon and text: "Limit number of gene sets and/or number of functions to avoid timeout. If you have large numbers of genes and/or functions, submit a computation job instead." Below the hint, there are two radio button options: "Use only functions in set:" with a dropdown menu showing "3_hydroxypropionate", and "Use all functions of type:" with a dropdown menu showing "COG". Below these, there is a label "MER-FS Metagenome:" followed by a dropdown menu showing "Assembled". A button labeled "Gene Set Function Profile" is positioned below the dropdowns. Below this button, there is a section titled "Submit as Computation Job Using Message System" with the text "You may submit a gene set function profile computation to run in the background." Below this text, there are two radio button options: "Save as a new job with name:" followed by a text input field, and "Replace the selected job:" followed by a dropdown menu showing "new1". A button labeled "Submit Computation" is at the bottom of the form.

Figure 5.5: Gene Set Function Profile

Gene Set Function Profile is similar to Genome Set Function Profile described in Section 4.2. Users can also submit profile as background computation job. Computation job is described in more detail in Section 7.

5.4 Set Operation

Set operation is similar to set operations on Workspace function sets described in Section 3.4.

6. Workspace Scaffold Sets

Workspace scaffold sets main page (Figure 5.1) lists all the scaffold sets created and owned by a user. If the user belongs to one or more IMG groups, then the table will also include scaffold sets shared by other group members, and the table will have two additional columns:

- Owner: the owner of scaffold set
- Shared with Group: all IMG groups that have access to this scaffold set

Scaffold sets sharing is similar to function sets sharing. Check Section 3 to find out how to share Workspace data sets.

My Workspace - Scaffold Sets

Scaffold Sets

Import & Export

Genomes & Genes

Function Profile

Histogram

Kmer Analysis

Phylogenetic Distribution

Set Operation

Add Selected to Scaffold Cart

Select All

Clear All

Remove Selected

Select	File Name	Number of Scaffolds (click the link to each individual set)	Owner	Shared with Group
<input type="checkbox"/>	16s	471	me	-
<input type="checkbox"/>	cleaned1	231	me	-
<input type="checkbox"/>	hawaii	85	me	-
<input type="checkbox"/>	hawaii5	5	me	-
<input type="checkbox"/>	jh_mixed_scaffolds_3	8	me	GBP
<input type="checkbox"/>	scaffold50	151	me	GBP
<input type="checkbox"/>	scaffold67	20	me	-
<input type="checkbox"/>	scaffold95	918	me	GBP
<input type="checkbox"/>	wetlandScaf	2	me	GBP
<input type="checkbox"/>	wetland_L1_cattail	51	me	GBP

(Note: You can only share or delete your own datasets.)

IMG Group: GBP

Share Selected Set(s) with Group

Remove Sharing of Selected Set(s)

Add Selected to Scaffold Cart

Select All

Clear All

Remove Selected

Figure 6.1: My Workspace - Scaffold Sets

Workspace scaffold set has 7 additional tabs with additional functions. We will describe these functions immediately below.

6.1 Import & Export

Import & Export is similar to Workspace function set Import & Export described in Section 3.1.

6.2 Genomes & Genes

To view genomes of scaffolds in select Workspace scaffold set(s), users

- select one or more Workspace scaffold sets;
- click the "View Genomes" button (in Figure 6.2).

View Genomes

You can view genomes of scaffolds in the selected scaffold set(s).

[View Genomes](#)

Save Genomes to My Workspace

hint:

Even though you can save large amount of data into workspace, many profile functions will timeout for extremely large workspace datasets

Save **all** genomes of selected scaffold set(s) to [My Workspace](#).
(Special characters in file name will be removed and spaces converted to _)

☒ Save to File name:
☐ Append to the following genome set:
☐ Replacing the following genome set:

[Save Genomes of Selected Scaffold Sets](#)

Figure 6.2: View Genomes and Save Genomes to My Workspace

Figure 6.3 shows an example result. Users can select genomes in the result table to load to Genome Cart or to save into a new Workspace genome set.

Genomes of Selected Scaffold Sets

Selected scaffold set(s): `jh_mixed_scaffolds_3`

Filter column: Domain Filter text: [Apply](#) [?](#)

[Export](#) Page 1 of 1 << first < prev 1 next > last >> All

Column Selector [Select Page](#) [Deselect Page](#)

Select	Domain	Status	Genome Name
<input type="checkbox"/>	*	D	Bankia setacea gill microbiome from Puget Sound WA, LANL assembly (Combined 454/Illumina v2 condensed contigs)
<input type="checkbox"/>	*	D	Hot spring thermophilic microbial communities from Obsidian Pool, Yellowstone National Park, USA - site 3 B9 (Yellowstone National Park Obsidian Pool site 3 B9, ASSEMBLY DATE=20131030)
<input type="checkbox"/>	A	F	Thermococcus gammatolerans EJ3

[Export](#) Page 1 of 1 << first < prev 1 next > last >> All

[Add Selected to Genome Cart](#)
[Select All](#)
[Clear All](#)

Save Genomes to My Workspace

hint:

Even though you can save large amount of data into workspace, many profile functions will timeout for extremely large workspace datasets

Save **selected** genomes to [My Workspace](#).
(Special characters in file name will be removed and spaces converted to _)

☒ Save to File name:

Figure 6.3: Genomes of Selected Scaffold Sets

Users can use the Save Genes to My Workspace function (Figure 6.4) to save all genes in scaffolds of selected Workspace scaffold set(s) into a new Workspace gene set. Since the number of genes is usually very large, we don't provide the view genes of scaffold sets function.

Save Genes to My Workspace

hint:

Even though you can save large amount of data into workspace, many profile functions will timeout for extremely large workspace datasets

Save **all genes** of selected scaffold set(s) to [My Workspace](#).
(Special characters in file name will be removed and spaces converted to _)

☒ Save to File name:

☐ Append to the following gene set:

☐ Replacing the following gene set:

cnb476

Save Genes of Selected Scaffold Sets

Figure 6.4: Save Genes to My Workspace

6.3 Function Profile

Users can select to perform Scaffold Set Function Profile based on (i) a selected Workspace function set or (ii) a function type (COG, Pfam, etc.). To perform Scaffold Set Function Profile, users

- select one or more scaffold sets;
- select one of the following (See Figure 6.5):
 - select a Workspace function set in the "Use only functions in set" dropdown list, or
 - select a function type in the "Use all functions of type" dropdown list;
- click "Scaffold Set Function Profile" button.

Scaffold Set Function Profile

hint:

Limit the number of scaffolds and/or number of functions to avoid timeout. If you have a large number of scaffolds and/or functions, please submit a computation job instead.

☒ Use only functions in set:

3_hydroxypropionate

☐ Use all functions of type:

COG

 MER-FS Metagenome:

Assembled

Scaffold Set Function Profile

Submit as Computation Job Using Message System

You may submit a scaffold set function profile computation to run in the background.

☒ Save as a new job with name:

☐ Replace the selected job:

test6

Submit Computation

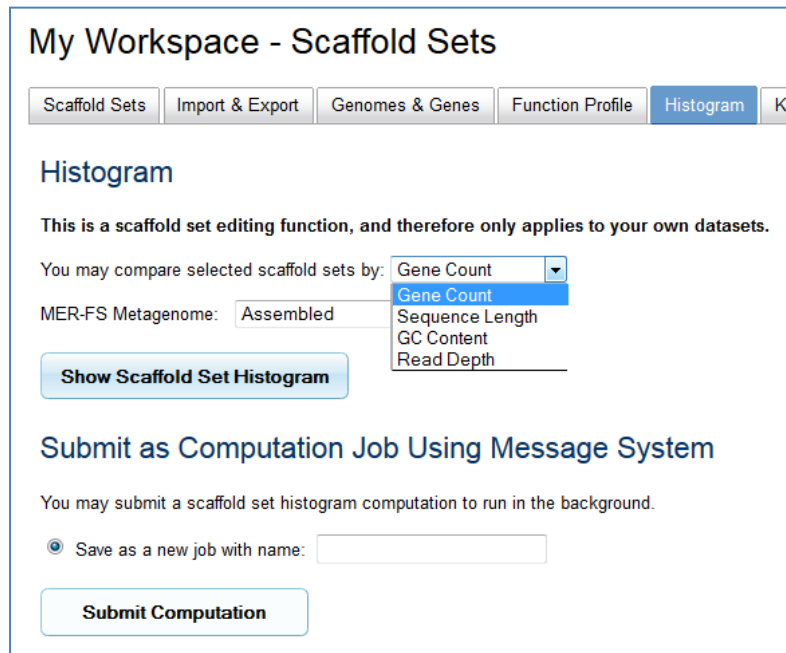
Figure 6.5: Scaffold Set Function Profile

Scaffold Set Function Profile is similar to Genome Set Function Profile described in Section 4.2. Users can also submit profile as background computation job. Computation job is described in more detail in Section 7.

6.4 Histogram

IMG supports 4 types of scaffold set histogram: (1) by Gene Count, (2) by Sequence Length, (3) by GC Content, and (4) by Read Depth. To view scaffold set histogram, users:

- select one or more scaffold sets;
- select a histogram type (Gene Count, Sequencing Length, etc.) in the dropdown list (in Figure 6.6);
- click the "Show Scaffold Set Histogram" button.



My Workspace - Scaffold Sets

Scaffold Sets Import & Export Genomes & Genes Function Profile **Histogram** Knowledge

Histogram

This is a scaffold set editing function, and therefore only applies to your own datasets.

You may compare selected scaffold sets by: Gene Count

MER-FS Metagenome:

Submit as Computation Job Using Message System

You may submit a scaffold set histogram computation to run in the background.

☒ Save as a new job with name:

Figure 6.6: Histogram

Figure 6.7 shows an example Scaffold Sets by Gene Count Histogram. Users can click on the bars in the graph to view more detailed scaffold information. The table display of the histogram is shown in the "Scaffolds by Gene Count" tab with clickable counts to detailed information. Again, users can select scaffolds in the detailed list to be loaded into Scaffold Cart or saved to a new Workspace scaffold set.

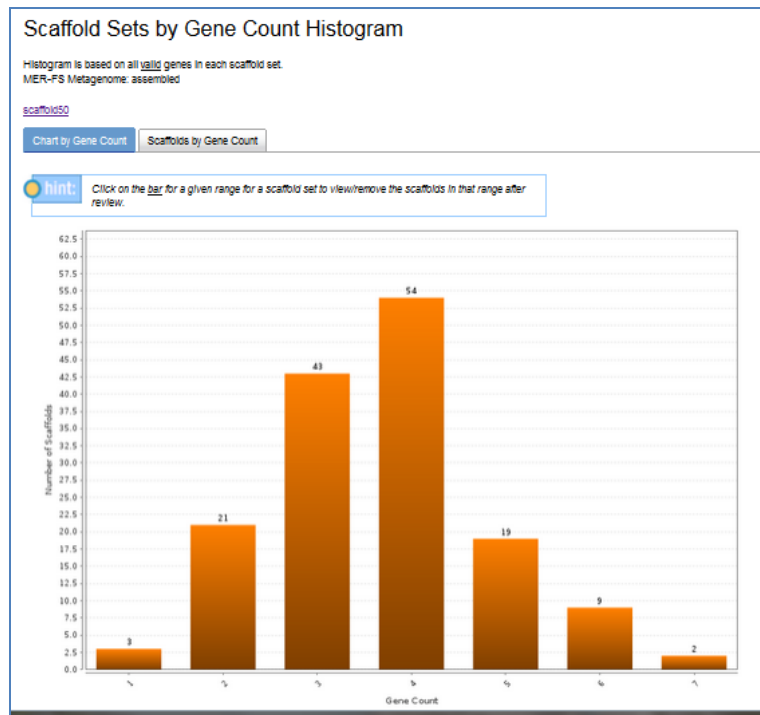


Figure 6.7: Scaffold Sets by Gene Count Histogram

6.5 Kmer Analysis

Kmer analysis provides 2D and 3D view of scaffold consistency check. To perform Kmer analysis, users:

- select one or more Workspace scaffold sets;
- select graphic display parameters (in Figure 6.8);
- click the "Kmer Frequency Analysis" button.

My Workspace - Scaffold Sets

Scaffold Sets
Import & Export
Genomes & Genes
Function Profile
Histogram
Kmer Analysis

Scaffold Consistency Check

You may analyze selected scaffold sets for purity using Kmer Frequency Analysis.

MER-FS Metagenome: Assembled

Kmer Frequency Analysis

Submit as Computation Job Using Message System

You may submit a scaffold set kmer computation to run in the background.

Lowering the 'Oligomer size' helps avoid memory issues

Parameter	Setting
Fragment window (1000 - 10000)	5000 bp
Fragment step (100 - 1000)	500 bp
Oligomer size (2 - 8)	4
Minimum variation (1 - 20)	10

☒ Save as a new job with name:

Submit Computation

Figure 6.8: Kmer Analysis

Figure 6.9 and Figure 6.10 show example Kmer analysis result in 2D view and 3D view, respectively.

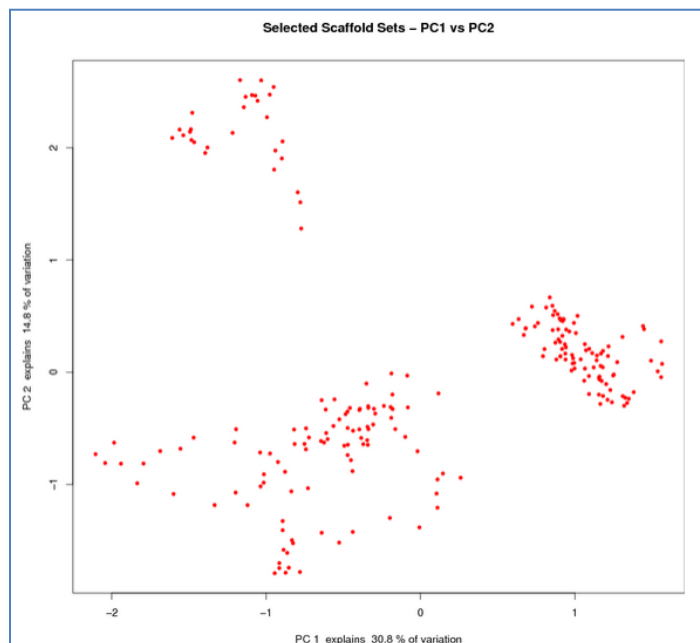


Figure 6.9: Kmer Analysis Result in 2D View

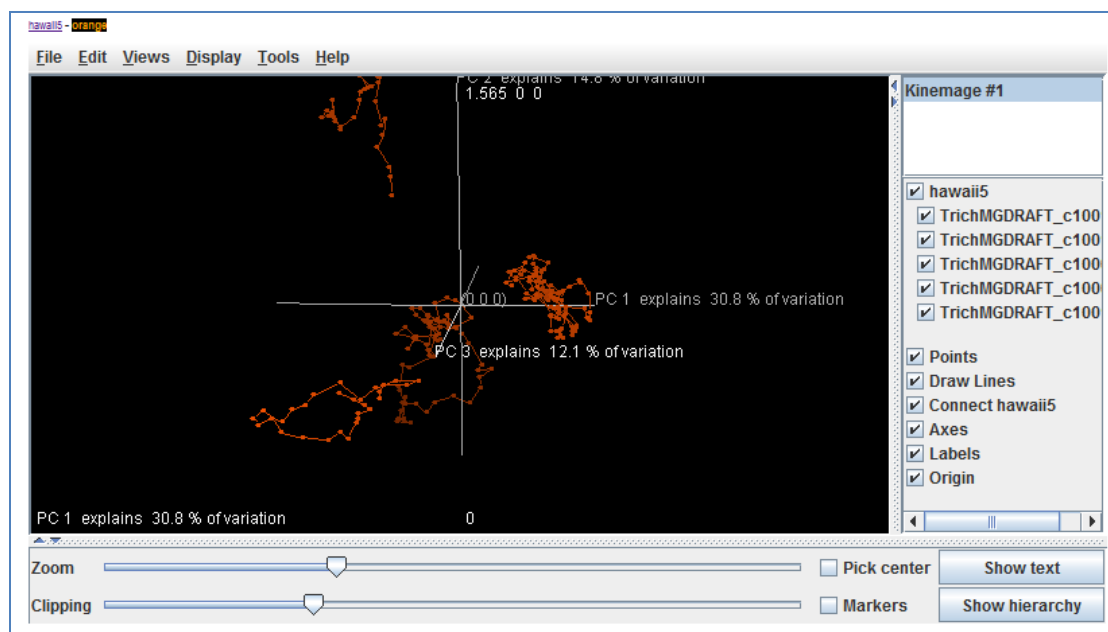


Figure 6.10: Kmer Analysis Result in 3D view

Both 2D view and 3D view are clickable to show more detailed information.

Users also have the option to submit Kmer analysis as computation jobs.

6.6 Phylogenetic Distribution

To perform phylogenetic distribution analysis of scaffolds, users:

- select one or more Workspace scaffold sets;
- click "Distribution by BLAST percent identities" button (in Figure 6.11).

Phylogenetic Distribution

You may view the phylogenetic distribution of best blast hits of protein-coding genes in selected scaffold sets.

MER-FS Metagenome: Assembled

Distribution by BLAST percent identities

Submit as Computation Job Using Message System

You may submit a scaffold set phylogenetic distribution computation to run in the background.

☒ Save as a new job with name:

☐ Replace the selected job: phylo1

Submit Computation

Figure 6.11: Scaffold Set Phylogenetic Distribution

Figure 6.12 shows an example result of scaffold set phylogenetic distribution. Users can click on Phylum field values or gene counts to view more detailed information.

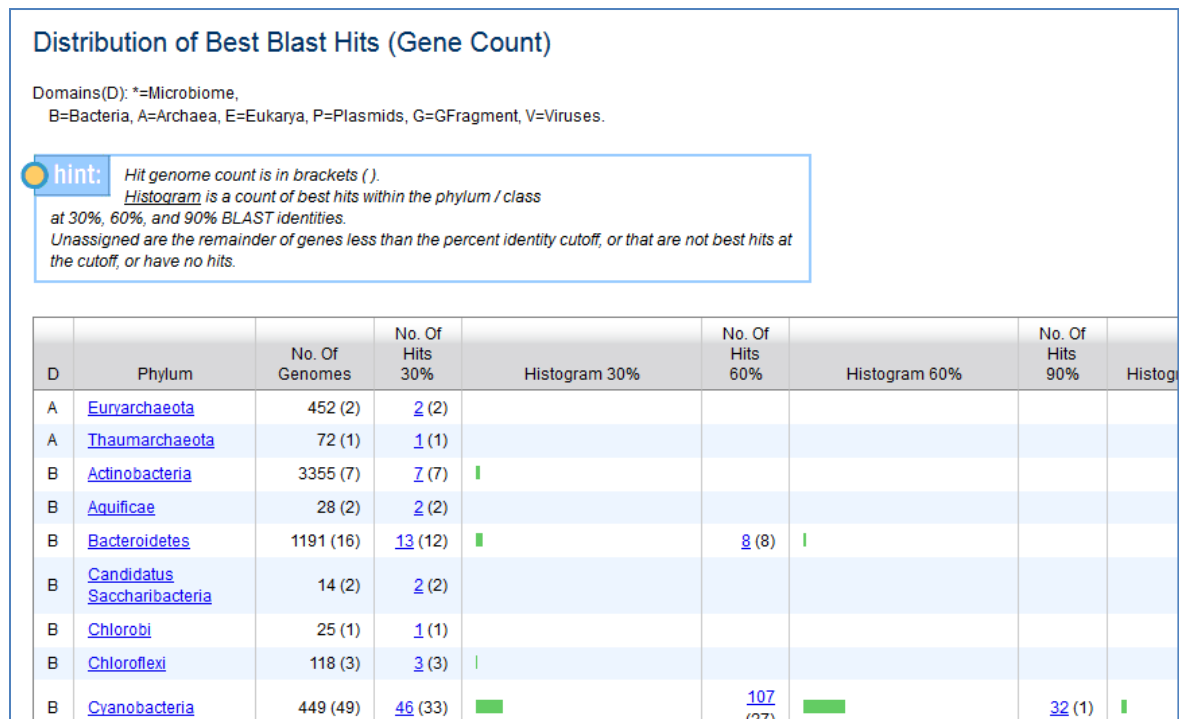


Figure 6.12: Phylogenetic Distribution of Genes in Selected Scaffold Sets

Users also have the option to submit phylogenetic distribution as computation jobs.

6.7 Set Operation

Set operation is similar to set operations on Workspace function sets described in Section 3.4.

7. Computation Jobs

IMG supports computation jobs using message system for users to submit long running jobs that will/may cause the Web UI to time out.

There are 2 ways to view your list of submitted jobs:

1. From MyIMG -> MyJob
2. From MyIMG -> Workspace, Computation Jobs

IMG User Computation Jobs

Computation Jobs Using Message System

You can view and track all your computation jobs using Message System.
(Click on the count to view the full list of jobs.)

Computation Jobs
11

Figure 7.1: IMG User Computation Jobs

Click on the count (in Figure 7.1) to view a complete job list:

Computation Jobs

Filter column: Name

Filter text

Apply

?

Export

Page 1 of 1

<< first < prev 1 next > last >>

All

Column Selector

Select Page

Deselect Page

Select	Name	Type	Start Time	Parameters	End Time	Status
<input type="checkbox"/>	test9	Scaffold Function Profile	8/29/2013 11:59:23	--function cog1to9 --scaffold hawaii	8/29/2013 11:59:52	completed
<input type="checkbox"/>	test8	Scaffold Phylo Distribution	4/30/2013 15:17:48	--scaffold scaffold31	4/30/2013 15:18:10	completed
<input type="checkbox"/>	test6	Scaffold Function Profile	10/10/2012 11:38:56	--functype COG --scaffold hawaii	10/10/2012 12:34:54	completed
<input type="checkbox"/>	test5	Scaffold Phylo Distribution	2/8/2013 11:8:30	--scaffold wetlandScaf,wetland_L1_cattail	2/8/2013 11:8:36	completed
<input type="checkbox"/>	test4	Gene Function Profile	11/14/2012 9:56:27	--function scog --gene etoliko1,etoliko2	11/14/2012 9:58:49	completed
<input type="checkbox"/>	test3	Scaffold Phylo Distribution	2/20/2013 14:4:50	--scaffold wetlandScaf,wetland_L1_cattail	2/20/2013 14:5:35	completed
<input type="checkbox"/>	test2	Gene Function Profile	11/12/2012 13:57:54	--functype Pfam --gene Human_Stool_and_Buccal_Genes_with_COG0002	11/12/2012 17:18:8	completed
<input type="checkbox"/>	test1	Gene Function Profile	11/12/2012 11:56:12	--functype TIGRfam --gene 6370000290essen,637000126essen,637000184essen	11/13/2012 12:9:39	completed
<input type="checkbox"/>	phylo1	Scaffold Phylo Distribution	5/9/2013 11:11:41	--scaffold mouse	5/9/2013 15:48:49	completed
<input type="checkbox"/>	new1	Gene Function	8/29/2013	--functype TIGRfam --gene 6370000290essen	8/29/2013	error (256)

Figure 7.2: Computation Jobs List

Each computation job has the following properties:

- Name: name of the job
- Type: job type, which can be Scaffold Function Profile, Gene Function Profile, etc.
- Start Time: the start time of the job
- Parameters: parameters associated with the job
- End Time: the end time of the job
- Status: the status of the job, which can be Completed, Error, or Processing

Click on the job status to view the detailed information and result (if completed) of the job. See Figure 7.3 for an example computation job report.

Computation Job: test9

Job Type: Scaffold Function Profile

--function cog1to9

--scaffold hawaii

8/29/2013 11:59:23

Status: Completed

Filter column: Function ID Filter text: Apply ?

Export

Page 1 of 1

<< first < prev

1

next > last >>

All

Column Selector

Select Page

Deselect Page

Select	Function ID	Function Name	hawaii
<input type="checkbox"/>	COG0003	Anion-transporting ATPase, ArsA/GET3 family	1

Export

Page 1 of 1

<< first < prev

1

next > last >>

All

Select All

Clear All

Save to My Workspace

Save selected functions or genes of selected functions to [My Workspace](#).

(Special characters in file name will be removed and spaces converted to _)

Figure 7.3: Example Computation Job Report